Puma Golf Ball Venture Analysis

Christian Farfan

Table of Contents **Page**

Project Background, Scope, and Development of Hypothesis

In this case study of Puma's new business venture, we are exploring the possibility of introducing a longer lasting, more cut resistant golf ball to the market. We will be comparing the driving distances in yards of the current ball that is on the market, and a new cut resistant ball. We are attempting to find out if there is a statistical difference between the driving distances of the current ball and the new ball. In order to achieve this, we will analyze the differences between two population means when both populations have matched samples. We have 40 driving samples for each ball.

We are trying to find $\boldsymbol{\mu_d = \mu_1 - \mu_2}$, where μ_1 is the population mean of the driving distances of the current model golf balls. μ_2 is the population mean of the driving distances of the new longer lasting, cut resistant golf balls. μ_d is the population mean of the difference.

The null hypothesis states that there is no significant difference in the driving distances of the current model golf balls and the new cut resistant golf balls.

- H_0: $\mu_d = 0$

The alternative hypothesis states that there is a significant difference between the current model golf balls and the new cut resistant golf balls.

- H_a: $\mu_d \neq 0$

If we **reject the null hypothesis**, we would accept the alternative hypothesis as true. This would conclude that there is a significant difference in the driving distances of the current model golf balls versus the new cut resistant model of the golf ball.

If we **do not reject the null hypothesis**, we are accepting the null as true. This would conclude that there was no significant difference in the driving distances of the current model golf balls versus the new cut resistant model of the golf ball.

If there is no significant difference in the driving distances of the balls, it would be advisable for Puma to conduct additional market research and perform comprehensive production and cost analysis to thoroughly assess the following factors:

- Are there additional costs to produce and manufacture the new ball compared to the current ball, and if so, are they significant to the point where the new ball is not a viable option?
- How much would consumers pay for the new ball compared to the current ball?
- Are competitors looking to introduce similar products?
- Would the new ball cannibalize the sale of the current ball, and if so, by how much?

If there is a significant difference between the driving distances of the balls, Puma may want to consider if the new ball travels further or shorter than the current ball, and if shorter, by how much. Puma would also need to perform market research and analysis to determine if consumers would pay for a ball that has a shorter driving range, but is more cut resistant and longer lasting.

Data Integrity

```
> str(sports_data)
tibble [40 × 2] (S3: tbl_df/tbl/data.frame)
 $ Current: num [1:40] 264 261 267 272 258 283 258 266 259 270 ...
 $ New    : num [1:40] 277 269 263 266 262 251 262 289 286 264 ...
```

Figure 1: Data Set Structure

To begin our analysis, we want to confirm that the data we imported is free of error, reliable, and complete. Figure 1 gives us an overview of the data set. We can see that our data set is considered a data frame. We also see some samples from each ball displayed, as well as "num", indicating that our columns are considered numeric. Figure 1 also displays the expected 40 samples for each ball, in this case titled "Current" and "New". We renamed the columns to specify what 'Current' and 'New' are.

```
> head(sports_data)
# A tibble: 6 × 2
  current_ball new_ball
         <dbl>    <dbl>
1          264      277
2          261      269
3          267      263
4          272      266
5          258      262
6          283      251
```

Figure 2: Columns Renaming

After renaming the columns, we want to ensure that our data has retained its integrity. The "Current" column was renamed to "current_ball", and the "New" column was renamed to "new_ball". From figure 2 we can also see that the first 6 samples for each ball are congruent with figure 1.

```
> # Checking for any null and/or missing values
> complete.cases(sports_data)
 [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[24] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
> anyNA(sports_data)
[1] FALSE
> is.na(sports_data$current_ball)
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[20] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[39] FALSE FALSE
> is.na(sports_data$new_ball)
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[20] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[39] FALSE FALSE
```

Figure 3: Missing Values Check

Figure 3 indicates that there are no missing values in our data set. We check the data set as a whole as well as the individual columns in order to cross-check our results.

Exploratory Data Analysis

```
> summary(sports_data)
  current_ball      new_ball
 Min.   :255.0   Min.   :250.0
 1st Qu.:263.0   1st Qu.:262.0
 Median :270.0   Median :265.0
 Mean   :270.3   Mean   :267.5
 3rd Qu.:275.2   3rd Qu.:274.5
 Max.   :289.0   Max.   :289.0
```

Figure 4: Data Set Summary

Figure 4 displays a summary of our data set. We can see that the median and the mean for the current ball are 270 and 270.3 respectively. Looking at the new ball, we see that the median and the mean are 265 and 267.5 respectively.

```
> range_current_ball
[1] 255 289
> range_new_ball
[1] 250 289
```

Figure 5: Current and New Ball Ranges

Figure 5 gives us a look at the ranges for the current ball and the new ball. The current ball has a lower range or minimum of 255 yards while the new ball has a lower range of 250 yards. Both balls have an upper range, or a maximum, of 289 yards.

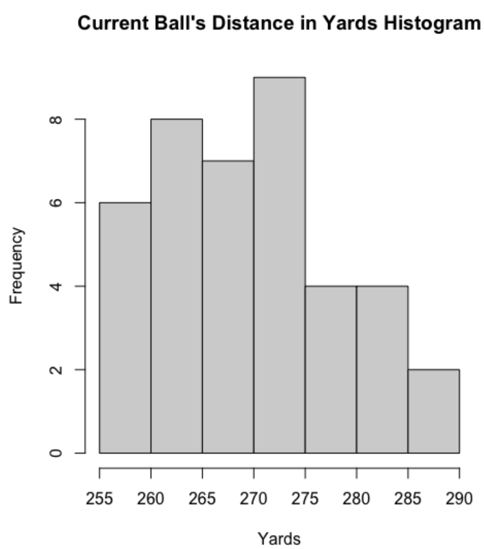

Figure 6: Current Ball Histogram

Figure 6 allows us to see the distribution of the driving distances in yards for the current ball. We can see that it is left skewed, with most of our samples falling between 255 yards and 275 yards.

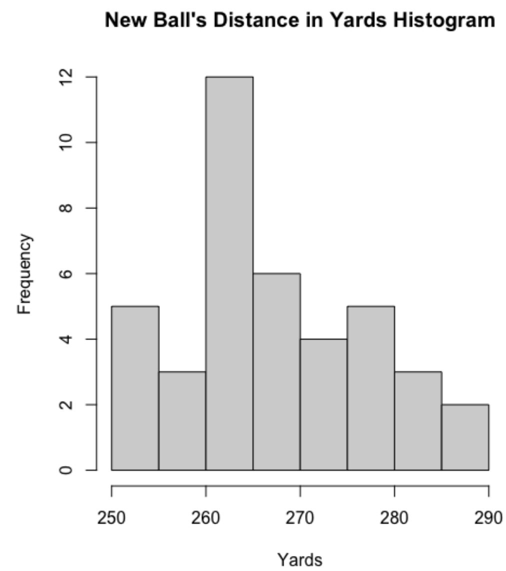

Figure 7: New Ball Histogram

Like figure 6, figure 7 allows us to see the distribution of the driving distances in yards for the new ball. While the histogram for the current ball was fairly ordinary with a left skew being the only notable trait, we can see that the histogram for the new ball is much more interesting. It is clearly not symmetric or uniformed, but rather has varying counts throughout its range.

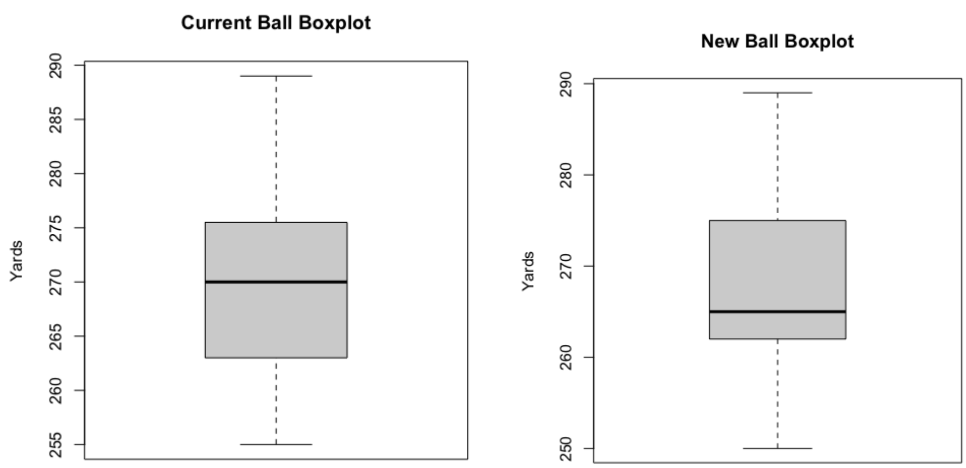

Figure 8: Current Ball and New Ball Boxplots

Continuing further with our exploratory analysis, we take a look at figure 8 which displays a boxplot for the current ball and the new ball. We see that there are no data points that are considered outliers for either ball.

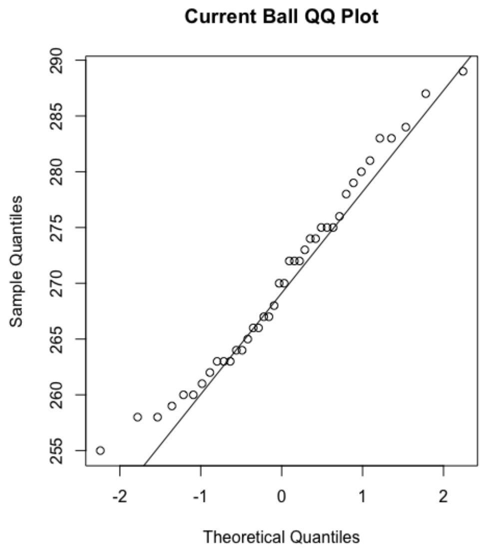

Figure 9: Current Ball QQ Plot

Figure 9 is a QQ plot of the current ball. From the distribution displayed, we can see that our data around the 260 yard mark and below deviates from the theoretical distribution. We can say that it has a slight left skew, which would be congruent with our figure 6 histogram. The QQ plot also displays some slight kurtosis, which indicates that we have more data points towards the center of our distribution.

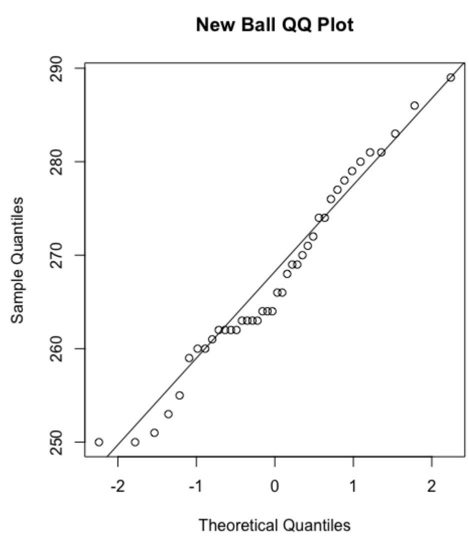

Figure 10: New Ball QQ Plot

The QQ plot for the new ball is more varied than the QQ plot of the current ball. Here we see that our data deviates around the mean and some of our lower values. The higher range of our plot seems to somewhat follow the expected theoretical distribution. The deviation of data from the lower range to the mean of our QQ plot can also be seen in the histogram of the new ball in figure 7.

New Data Frame with Differences Data

In order to continue with our statistical analysis, we take the difference of the current ball and the new ball. Before we did that, we pulled the columns for the current ball and the new ball from the data frame and created new variables for each column. We then created a new data frame with this new calculation. We also renamed the data frame. The figures below illustrate the process.

```
> # Pulling first column as Current data
> current_data <- sports_data[,1]
> current_data
# A tibble: 40 × 1
   current_ball
          <dbl>
 1          264
 2          261
 3          267
 4          272
 5          258
```

```
> # Pulling second column as New data
> new_data <- sports_data[,2]
> new_data
# A tibble: 40 × 1
   new_ball
      <dbl>
 1      277
 2      269
 3      263
 4      266
 5      262
```

Figure 11: Pulling Current and New Ball Columns

Figure 11 demonstrates the "current_ball" and "new_ball" columns being pulled and assigned to the variables "current_data" and "new_data" respectively.

```
> # Taking the difference between the current ball and the old ball
> differences_data = (current_data - new_data)
> differences_data
  current_ball
1          -13
2           -8
3            4
4            6
5           -4
```

Figure 12: Taking the Difference of Current and New Data

```
> differences_df
  current_ball new_ball current_ball.1
1          264      277            -13
2          261      269             -8
3          267      263              4
4          272      266              6
5          258      262             -4
```

```
> differences_df
  current_ball new_ball difference
1          264      277        -13
2          261      269         -8
3          267      263          4
4          272      266          6
5          258      262         -4
```

Figure 13: Creating New Data Frame and Renaming it

In figure 13 we create a new data frame where we incorporate the "differences_data" shown in figure 12, and rename it from "current_ball.1" to "difference".

Test Variables, Critical Value Method Test, and P-Value Method Test

In order to continue with our test, we find the mean and the standard deviation of our difference data, as well as set up and create our variables such as alpha and degrees of freedom.

```
> # Mean of difference data
> mean_differences <- mean(differences_df$difference)
> mean_differences
[1] 2.775
```

```
> # Sample standard deviation of difference data
> sample_sd_differences <- sd(differences_df$difference)
> sample_sd_differences
[1] 13.74397
```

Figure 13: Mean and Standard Deviation of Differences Data

```
> alpha = 0.05
> sample_size = 40
> degrees_of_freedom = 40 - 1
> alpha
[1] 0.05
> sample_size
[1] 40
> degrees_of_freedom
[1] 39
> degrees_of_freedom
[1] 39
```

Figure 14: Creating Variables for Alpha, Sample Size, and Degrees of Freedom

```
> # Creating our test statistic
> test_statistic = (mean_differences - 0) / (sample_sd_differences/sqrt(sample_size))
> test_statistic # The test statistic is 1.277.
[1] 1.27697
```

Figure 15: Calculating Test Statistic

The test statistic for our test is 1.277. We can reject the null hypothesis if the test statistic of 1.277 is less than or equal to $-t_{\alpha,df}$, or if the test statistic of 1.277 is greater than or equal to $-t_{\alpha,df}$. We calculated that $t_{0.05,39}$ is 2.023, and $-t_{0.05,39}$ is -2.023. The test statistic of 1.277 is not less than or equal to -2.023. The test statistic of 1.277 is not greater than or equal to 2.023.

```
> # Critical Value Method
> # Reject H0 if: test_statistic <= -t(alpha/2, degrees_of_freedom)
> #   OR t >= t(alpha/2, degrees_of_freedom)
> if (test_statistic <= -t_critical_value | test_statistic >= t_critical_value) {
+   print("Reject H0")
+ } else {
+   print("Do Not Reject H0")
+ }
[1] "Do Not Reject H0"
```

Figure 17: Critical Value T-Test Method Results

Therefore, using the critical value method, we conclude that we do not reject the null hypothesis.

```
> # Identifying two-tailed p-value
> p_value = 2*pt(test_statistic, degrees_of_freedom, lower.tail=FALSE)
> p_value # The p-value is 0.209.
[1] 0.2091636
```

Figure 18: Calculating P-Value

The p-value for our test is 0.209. We can reject the null hypothesis if the p-value of 0.209 is less than or equal to α, which is 0.05. Because 0.209 > 0.05, we do not reject the null hypothesis.

```
> # P-value method
> # Reject H0 if: p-value <= alpha (which is 0.05)
> if (p_value <= alpha) {
+   print("Reject H0")
+ } else {
+   print("Do Not Reject H0")
+ }
[1] "Do Not Reject H0"
```

Figure 19: P-Value Method Results

Therefore, using the p-value method, we conclude that we do not reject the null hypothesis. By not rejecting the null hypothesis, we are accepting that the null hypothesis is true.

By accepting the null hypothesis as true, we are rejecting the alternative hypothesis that there is a significant difference in the driving distances of the current-model golf balls and the new cut resistant golf balls. We utilized two statistical methods to cross-check our results and ensure accuracy.

The given sample sizes of the two types of golf balls (current and new) are greater than 30. The sample sizes are sufficient to test the driving distance of the two golf balls.

We can determine with 95% confidence that the driving distance for the current golf ball model ranges from 267.56 and 272.99 yards, and the new model of golf balls ranges from 264.43 and 270.57 yards (see Figure 13 below). We determined the lower and upper bounds by calculating the standard error and margin of error for both the new and current golf balls.

Additionally, we can determine with 95% confidence that the differences between the driving distances for the current and new model of golf balls ranges from -1.48 and 7.03 yards. We concluded this by using the standard error of the mean differences that we calculated earlier in our R program.

	Mean	Standard Deviation	Margin of Error	Lower Bound	Upper Bound
Current Model	270.275	8.753	2.713	267.56 yards	272.99 yards
New Model	267.5	9.897	3.067	264.43 yards	270.57 yards
Differences	2.775	13.744	4.259	-1.48 yards	7.03 yards

Figure 13: Descriptive Statistics Table

Table of Values

P-value	.209
Test statistic	1.277
$-t_{\alpha,df}$	-2.023
$t_{\alpha,df}$	2.023
α	0.05
Degrees of freedom	39
Lower bound (current)	267.562
Lower bound differences	-1.484
Lower bound (new)	264.433
Upper bound (current)	272.988
Upper bound differences	7.034
Upper bound (new)	270.567
Margin of error (current)	2.713
Margin of error differences	4.259
Margin of error (new)	3.067
Mean (current)	270.275
Mean differences	2.775
Mean (new)	267.5
Point estimate (current)	270.275
Point estimate differences	2.775
Point estimate (new)	267.5
Standard deviation (current)	8.753
Standard deviation (new)	9.900
Standard error (current)	1.384
Standard error (new)	1.565
z-value	1.960

Conclusion and Recommendation

We will provide a final summary of the case study and its business implications before presenting our conclusion and a possible recommendation.

Puma is considering introducing a longer lasting, cut resistant golf ball to the market. Before we can proceed with this business proposal, we need data to guide our decision. Our goal in this case study was to find out if there is a significant statistical difference between the driving distance of the new ball and the current ball.

In order to achieve this goal we analyzed two populations where both populations have an equal number of samples, in this case 40 samples for each ball. We conducted a critical value t-test as well as a p-value test and concluded that there was not a significant difference between the driving distances of the current ball and the new ball. With this result, we can proceed with the consideration of introducing the new cut resistant ball to the market.

Some factors to consider before approving the production of the new ball were mentioned earlier but they will be mentioned once more due to their crucial significance.

It is recommended for Puma to look at the landscape of the current and projected market in order to forecast the suitability of the new ball in the market. How will the new ball be received, how will it be marketed, is this a product that consumers in the industry want or need? Puma is also advised to consider the possibility of the new ball cannibalizing the sales of the current ball. If this does occur, what is the projected percentage of the sales that will be cannibalized from the current ball? Is this desired or will it put the company in a difficult position?

Puma should also consider performing production and cost analysis to determine once more if the new ball is a viable product. This research should also seek to find out if customers are willing to pay a premium for a longer lasting ball that travels as far as the previous product.

To conclude, there is no significant difference between the driving distance of the current model golf balls and the new cut resistant golf balls. We recommend that Puma expand their potential market share if the production, cost, and market research indicate that the new ball is a viable option.